Advances and Challenges in Computational Fusion Energy Science

Geophysical Fluid Dynamics Laboratory

Princeton, NJ

February 8, 2007

William M. Tang

Princeton University

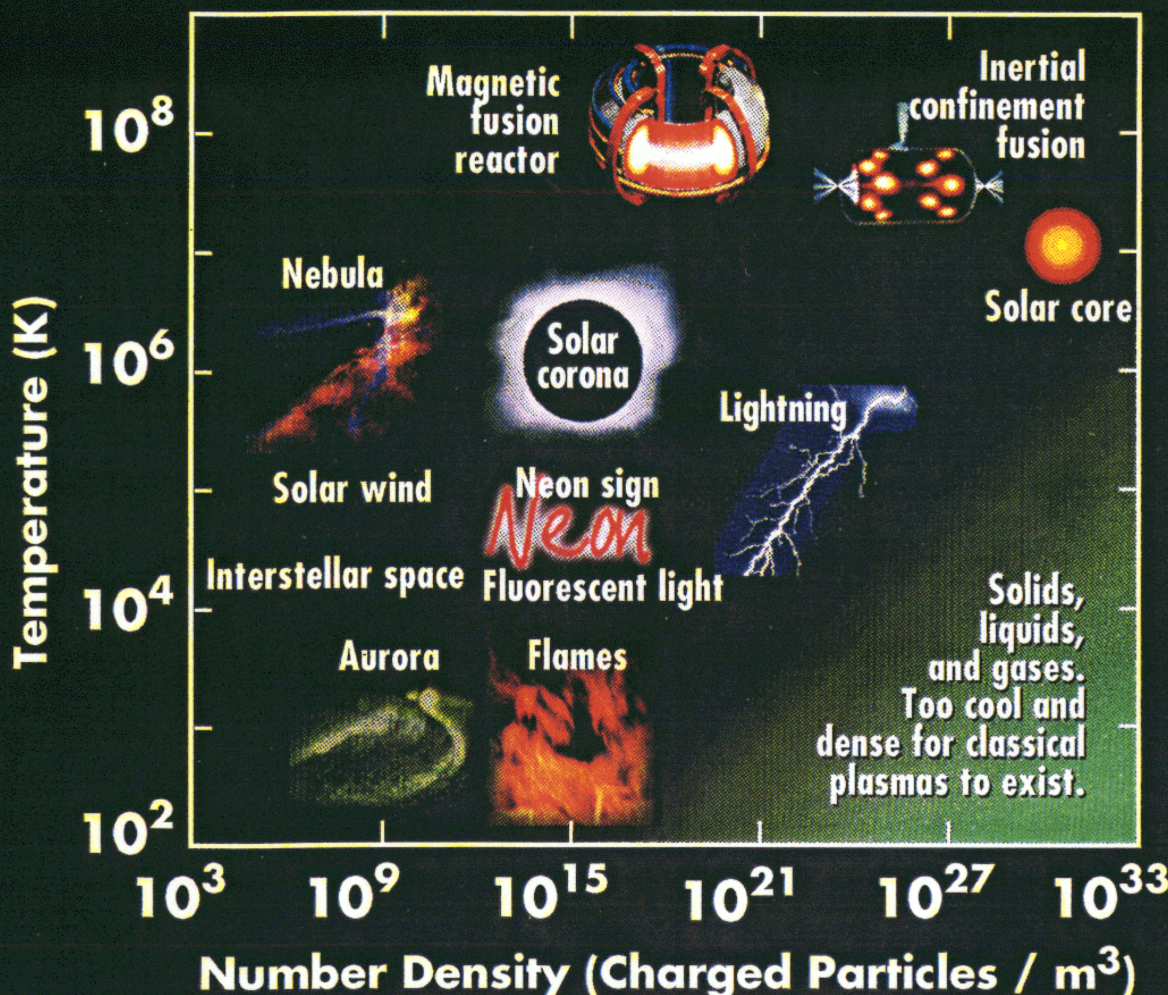
Chief Scientist, Princeton Plasma Physics Laboratory (PPPL)
and Associate Director, Princeton Institute for Computational Science and Engineering (PICSciE)

ADVANCED COMPUTING IS AN INCREASINGLY POWERFUL TOOL FOR SCIENTIFIC DISCOVERY

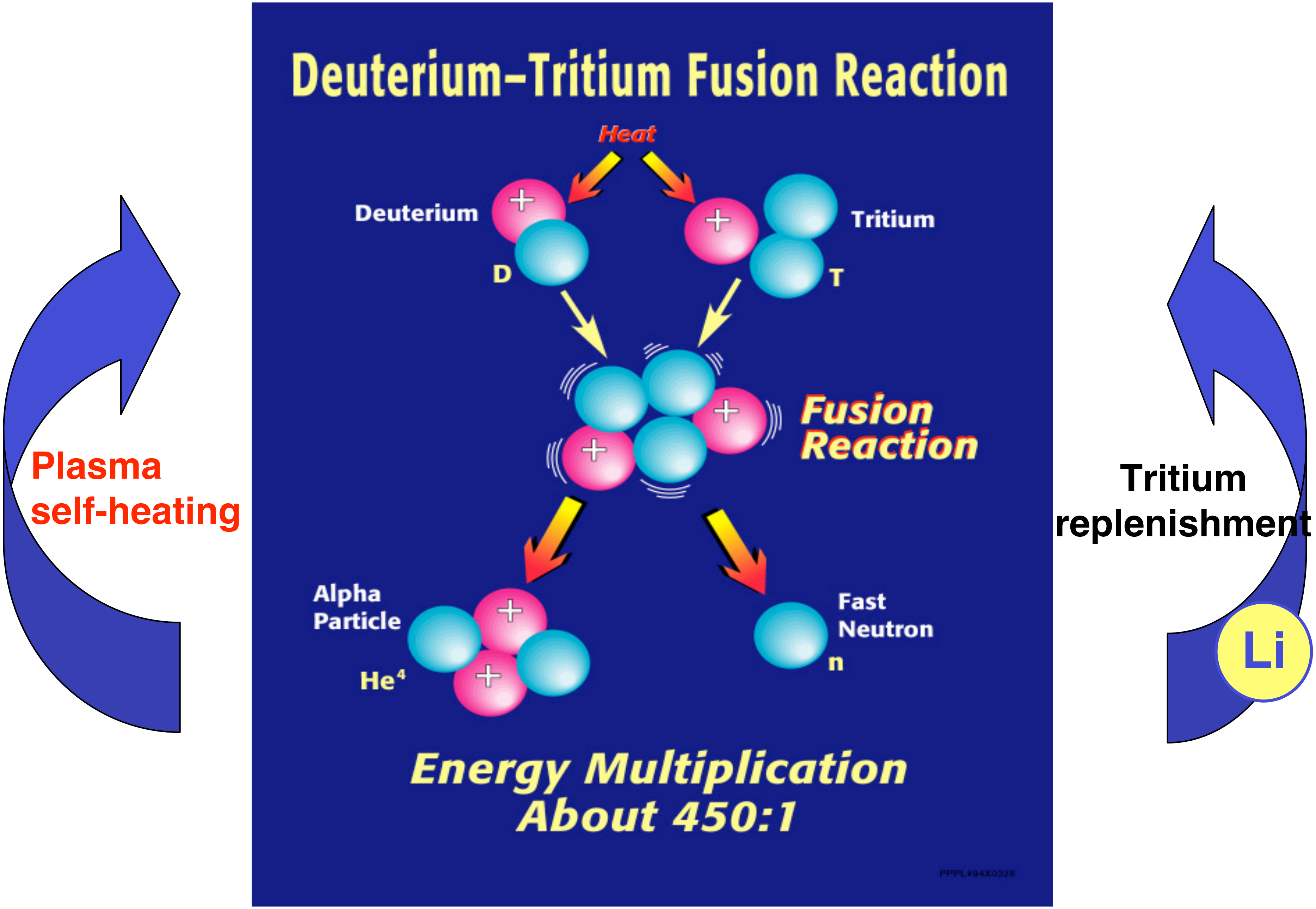
- **Advanced computation in tandem with theory and experiment has proven to be a powerful *new tool for scientific understanding and innovation* in many areas of research**

- **Plasma Science is *effectively utilizing* the exciting advances in Information Technology and Scientific Computing**
  - ***Reference*: *Advances and Challenges in Computational Plasma Science* Plasma Physics & Controlled Fusion 47 (February, 2005)**
  - **Accelerates progress toward reliable predictions of complex properties of high temperature fusion plasmas**
    - **Acquisition of *scientific understanding* needed for predictive models *superior to empirical scaling***
    - **Such models essential for *effective harvesting of knowledge* from present & future facilities such as ITER -- $10B international burning plasma project**

PLASMAS - THE 4TH STATE OF MATTER



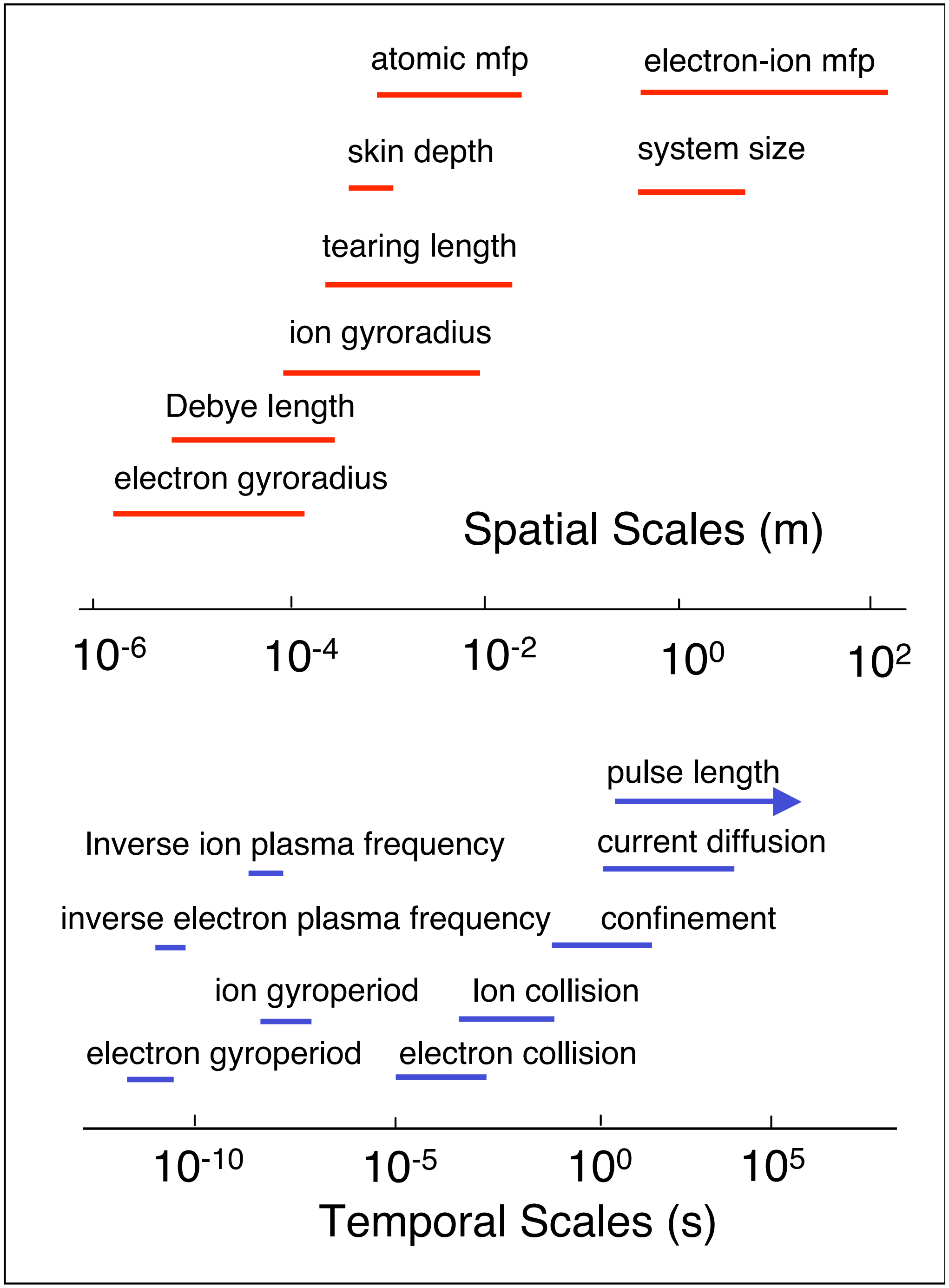
Fusion Energy Process



Spatial & Temporal Scales Present Major Challenge to Theory & Simulations

- **Huge range of spatial and temporal scales**

- **Overlap in scales often means strong (simplified) ordering not possible**



Plasma Physics Challenges

National Academy of Sciences Plasma Science Committee

Macroscopic Stability

***Fusion*: What limits the pressure in plasmas?**

***Solar Physics*: Solar flares**

Wave-particle Interactions

***Fusion*: How do hot particles and plasma waves interact in the nonlinear regime?**

***Space Physics*: Magnetospheric heating**

Microturbulence & Transport

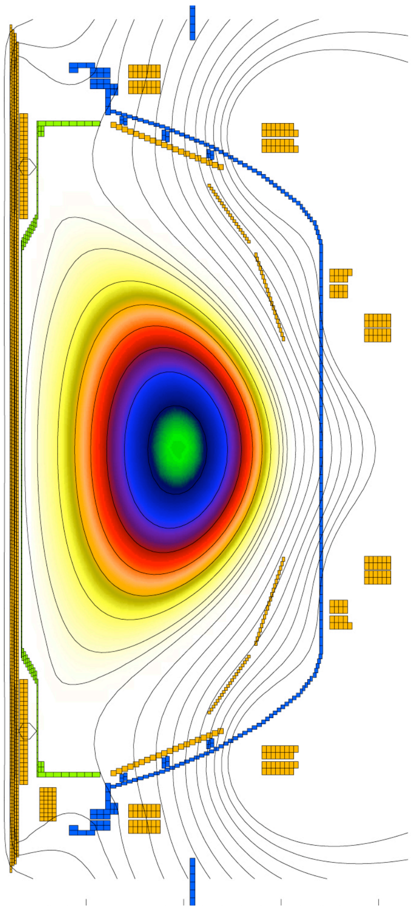
***Fusion*: What causes plasma transport?**

***Astrophysics*: Accretion disks (black holes)**

Plasma-material Interactions

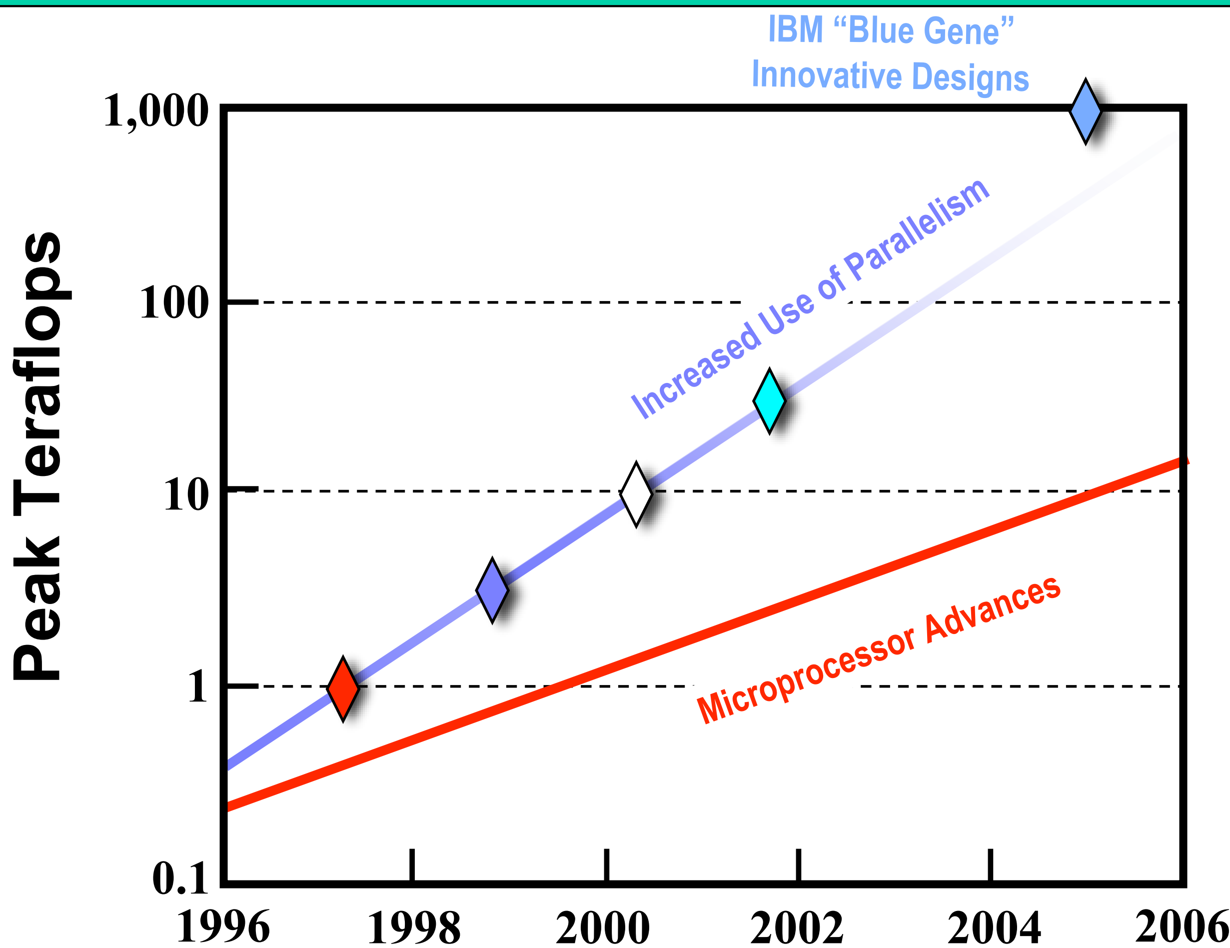
***Fusion*: How can high-temperature plasma and material surfaces co-exist?**

***Material Science*: Micro-electronics processing**



SciDAC
Scientific Discovery through Advanced Computing

Dramatic Advances in Computing: *Terascale Today, Petascale Tomorrow*



MICROPROCESSORS

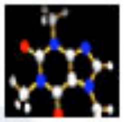
2x increase in microprocessor speeds every 18-24 months ("Moore's Law")

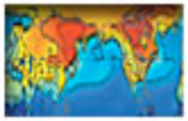
PARALLELISM

Many more processors being used on single problem

INNOVATIVE DESIGNS

Multi-core Processors




SciDAC Goal: Creation of 21st Century Computing Infrastructure built on "real" interdisciplinary collaborations

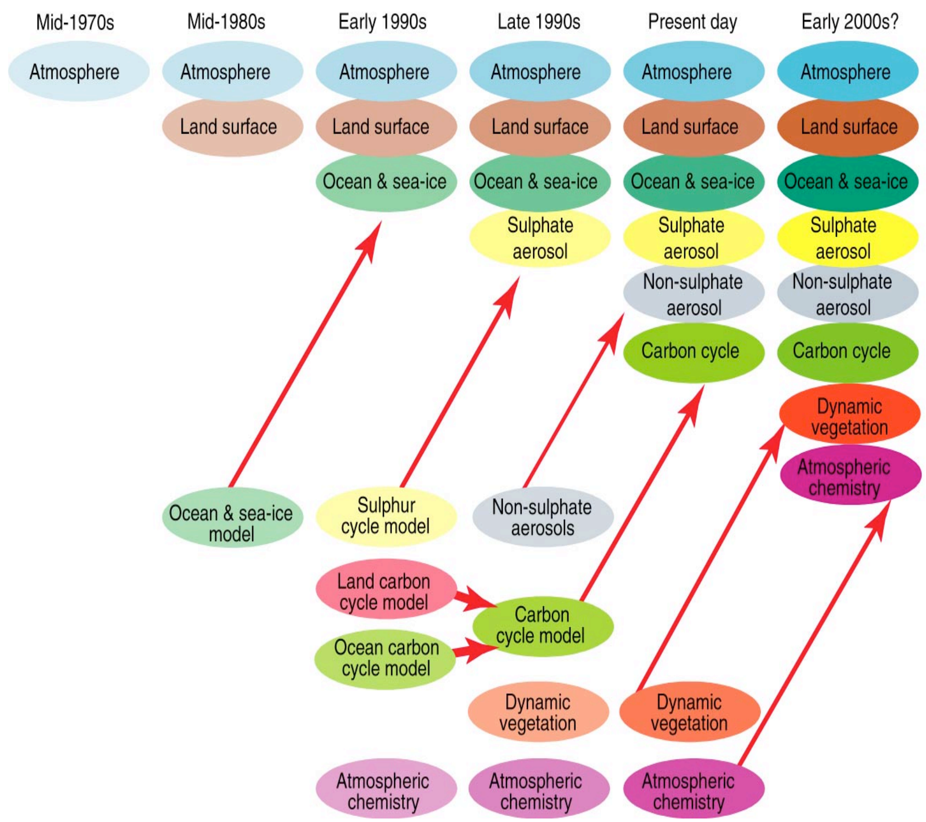
- ***Codes:*** new scientific domain applications codes capable of taking full advantage of terascale (and eventually petascale) computers
- ***Software Tools:*** new mathematical algorithms & solvers together with advanced systems operations capabilities to achieve maximum efficiency on HPC platforms
- ***Data Analysis & Management Methods:*** new data analysis methodologies with advanced visualization for knowledge extraction and management of unprecedented growth in huge data sets
- ***Networks:*** new networking technologies & collaboration tools needed to link geographically separated researchers

SciDAC Contributions to Community Model Building

Climate Collaborations: (J. Drake, ORNL)

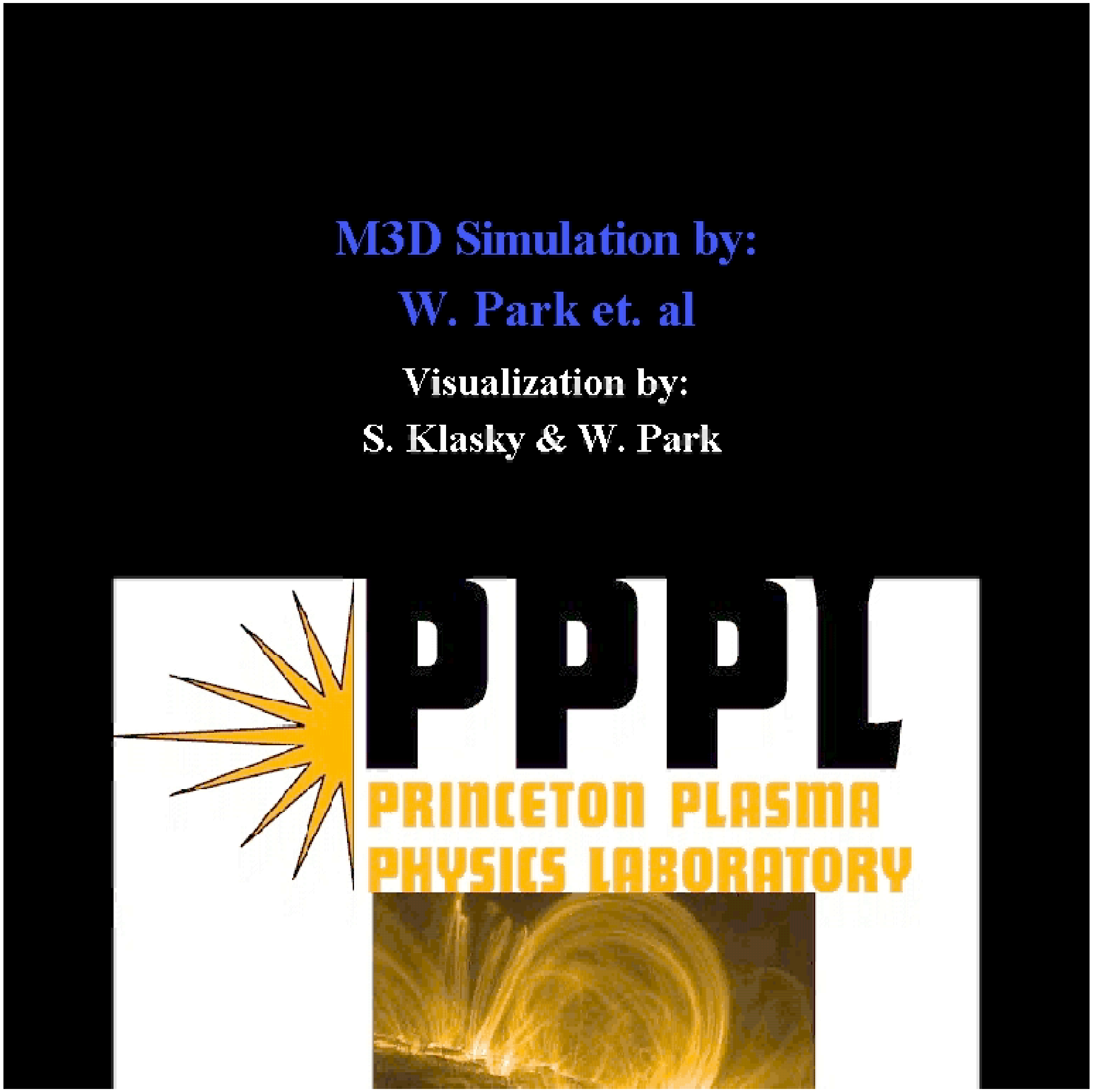
• Earth System Grid distributed the IPCC simulation data for the international community
• Performance studies with SciDAC PERC significantly increased throughput of CCSM
• Software engineering with CCA and the NASA Earth System Modeling Framework (ESMF) resulted in Model Coupling Toolkit (MCT) and conservative regridding packages
• Parallel I/O (pNetCDF) and analysis with SciDAC SDM enabled us to scale and develop high resolution models
• Mesh and operators with SciDAC TSTT forced generalization of our interfaces
• New formulations and consideration of "hard" problems with mathematicians in SciDAC APDEC

The Development of Climate models, Past, Present and Future

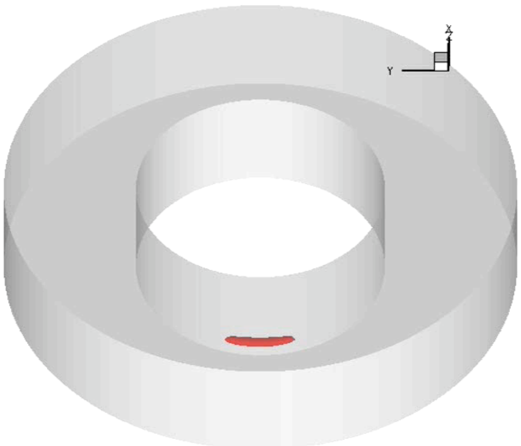


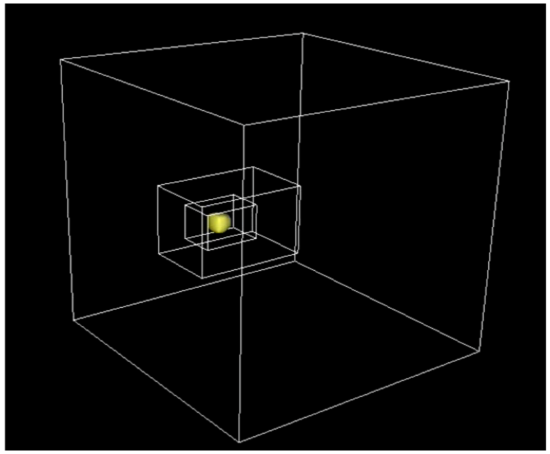
Community Model Building: *The NCAR Community Climate System Model (CCSM) has been assembled from components developed among agencies and universities with an open and peer reviewable process. This continues toward the development of a comprehensive earth system model.*
See J. Climate, vol 11, no 6 (1998) and vol 19, no 11 (2006).

Nonlinear MHD Advanced Simulation



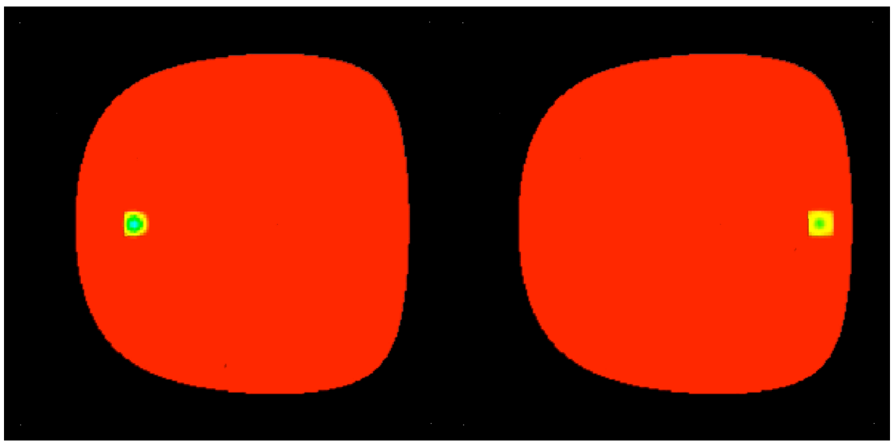
Fusion Codes Take Advantage of Latest Computational Advances *(R. Samtaney)*





Adaptive Mesh Refinement

Inside Pellet Launch →



← **Outside Pellet Launch**

Microturbulence in Fusion Plasmas

- Primary mechanism for cross-field transport in magnetically confined plasmas
  - Size and cost of a fusion reactor determined by balance between particle and energy confinement and fusion self-heating rates

- Challenge: complex multi-scale nonlinear problem
  - Large time and spatial scale separations similar to fluid turbulence (CFD)
  - Self-consistent accounting for electromagnetic fields: many-body problem
  - Strong nonlinear wave-particle interactions: kinetic dynamics
  - Microinstabilities driving turbulence require realistic representation of spatial inhomogeneities together with complex confining EM fields

Particle Simulation of the Boltzmann-Maxwell System

- The Boltzmann equation (*Nonlinear PDE in Lagrangian coordinates*):

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + \mathbf{v}\cdot\frac{\partial F}{\partial \mathbf{x}} + \left(\mathbf{E} + \frac{1}{c}\mathbf{v}\times\mathbf{B}\right)\cdot\frac{\partial F}{\partial \mathbf{v}} = C(F).$$

- "Particle Pushing" (*Linear ODE's*)

$$\frac{d\mathbf{x}_j}{dt} = \mathbf{v}_j, \qquad \frac{d\mathbf{v}_j}{dt} = \frac{q}{m}\left(\mathbf{E} + \frac{1}{c}\mathbf{v}_j \times \mathbf{B}\right)_{\mathbf{x}_j}.$$

- Klimontovich-Dupree representation,

$$F = \sum_{j=1}^{N}\delta(\mathbf{x}-\mathbf{x}_j)\delta(\mathbf{v}-\mathbf{v}_j),$$

- Poisson's Equation: [*Linear PDE in Eulerian coordinates (lab frame)*]

$$\nabla^2\phi = -4\pi\sum_{\alpha} q_\alpha \sum_{j=1}^{N}\delta(\mathbf{x}-\mathbf{x}_{\alpha j})$$

- Ampere's Law and Faraday's Law [*Linear PDE's in Eulerian coordinates (lab frame)*]

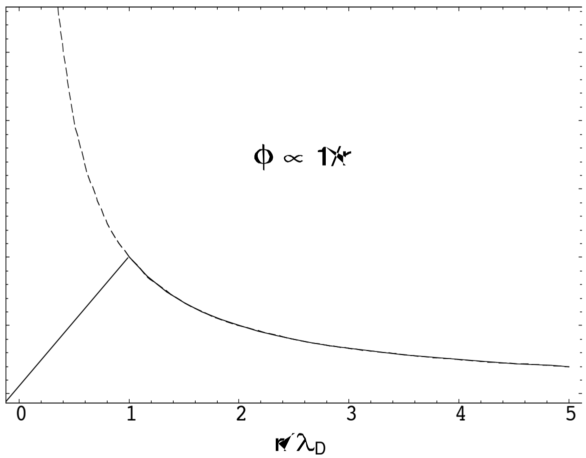
Particle-in-Cell Simulations

- **Early attempts [*Buneman (1959); Dawson (1962)*]**

- **Finite-Size Particles and Particle-in-Cell Simulation [*Dawson et al. (1968) and Birdsall et al. (1968)*]**

  - **Coulomb potential is modified for a finite size particle due to Debye shielding**

  - **no need to satisfy $1/(n\ \lambda_D^3) \ll 1$**



- **Number of calculations for N particles**

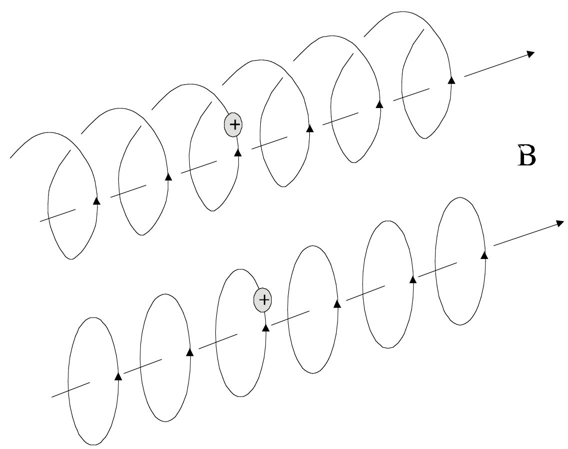
  - **N^2 for direct interactions and N for PIC**

- **Collisions are treated as sub-grid phenomena via Monte-Carlo methods [*Shanny, Dawson & Greene (1976)*]**

Gyrokinetic Particle Simulation

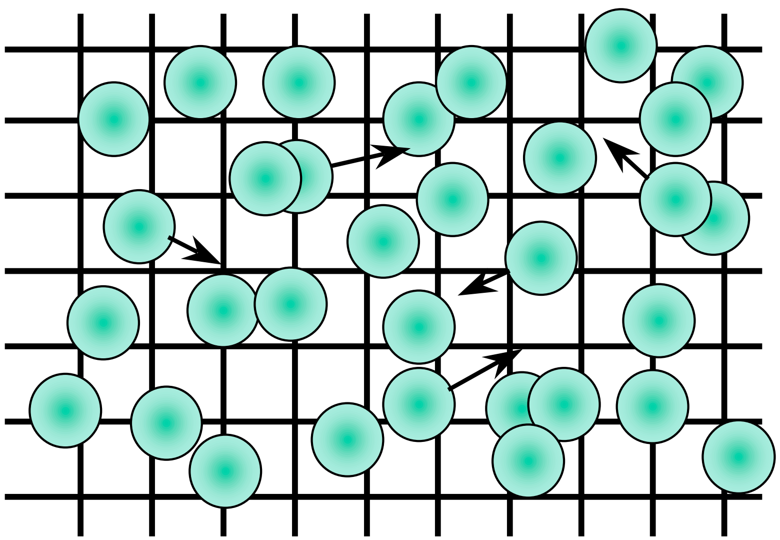
- *[W. Lee, PF ('83); JCP ('87)]*

- **Gyrophase-averaged Vlasov-Maxwell equations for low frequency microinstabilities.**
- **Spiral motion of a charged particle is modified as a rotating charged ring subject to guiding center electric and magnetic drift motion as well as parallel acceleration --** ***speeds up computations*** **by 3 to 6 orders of magnitude in time steps and 2 to 3 orders in spatial resolution**



Particle-in-cell (PIC) Method

- Particles sample distribution function (markers).
- The particles interact via a grid, on which the potential is calculated from deposited charges.

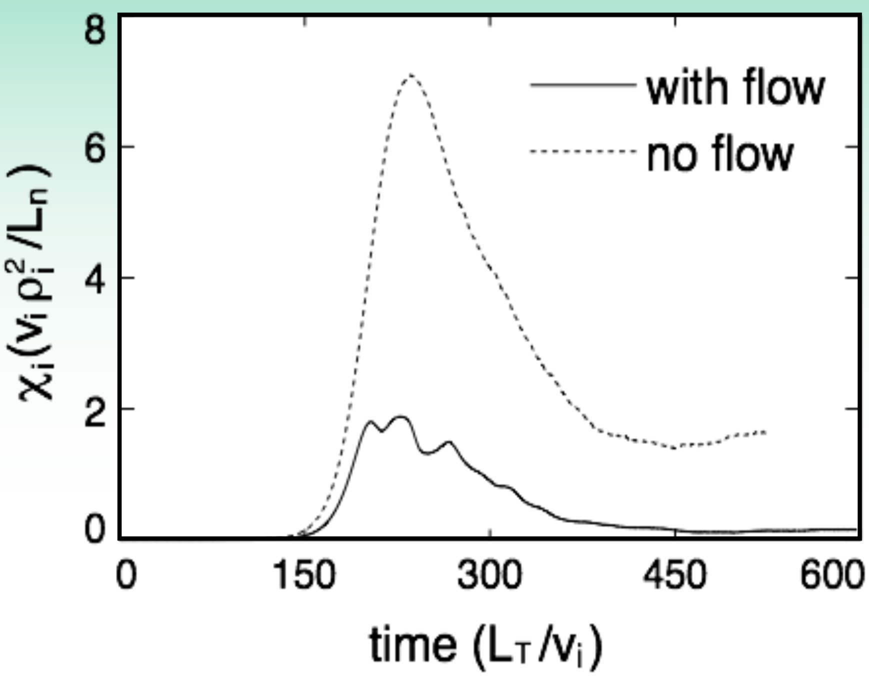


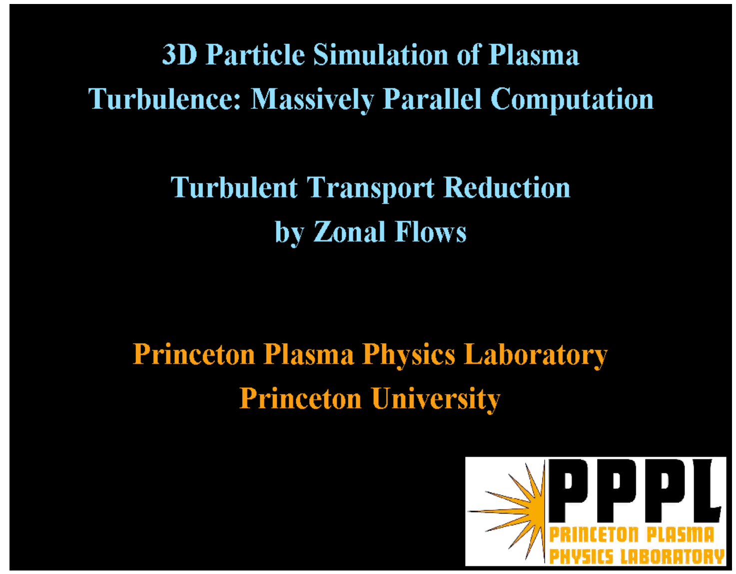
The PIC Steps

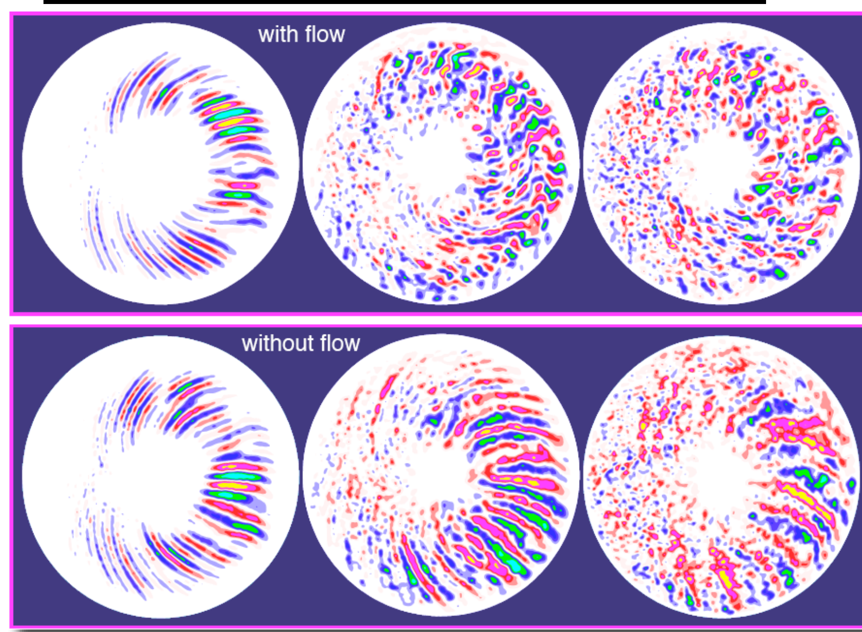
- **"SCATTER", or deposit, charges on the grid (nearest neighbors)**
- **Solve Poisson equation**
- **"GATHER" forces on each particle from potential**
- **Move particles (PUSH)**
- **Repeat…**

1998: Nonlinear zonal flow simulations by GTC with 10^8 particles on Cray T3E

Nonlinearly generated zonal flows (associated with ITG turbulence) break up the eddies and reduce transport in global simulations [Lin, et al, Science 1998]







3D Particle Simulation of Plasma Turbulence: Massively Parallel Computation

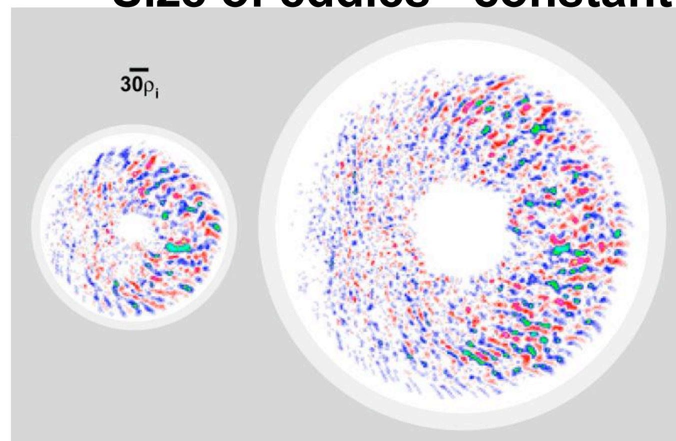
Turbulent Transport Reduction by Zonal Flows

Princeton Plasma Physics Laboratory
Princeton University

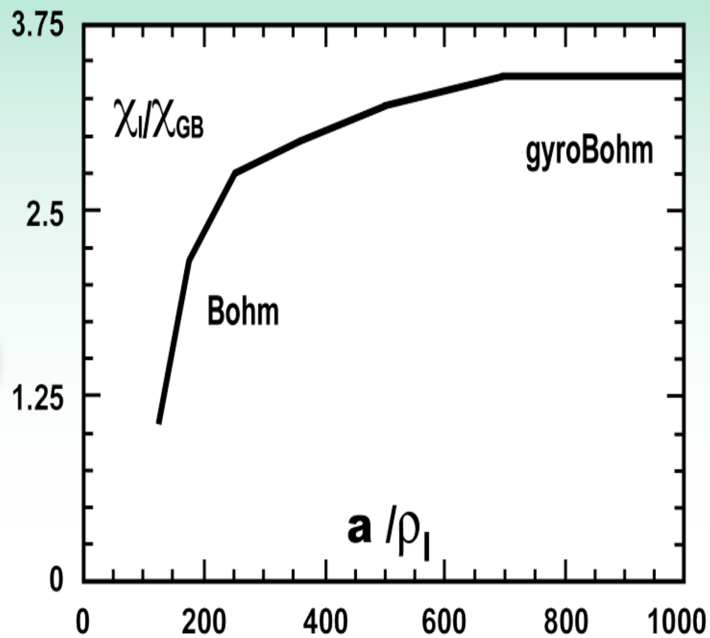
First Global ITER-size Simulation (2002) using 10^9 particles on IBM SP 3

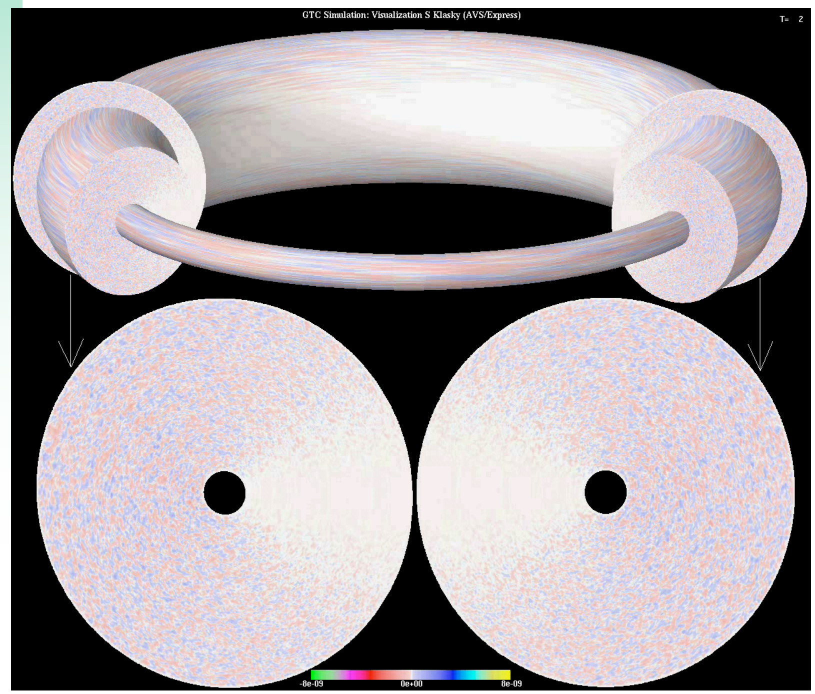
- *"Scientific Discovery"* - Transition to favorable scaling of confinement observed for large plasmas of future [Lin, et al., *PRL2002*]
- *Data Streaming Technology* enabled moving terabytes of data from NERSC to PPPL [Klasky, et al., *SC2003*]

Size of eddies ~constant



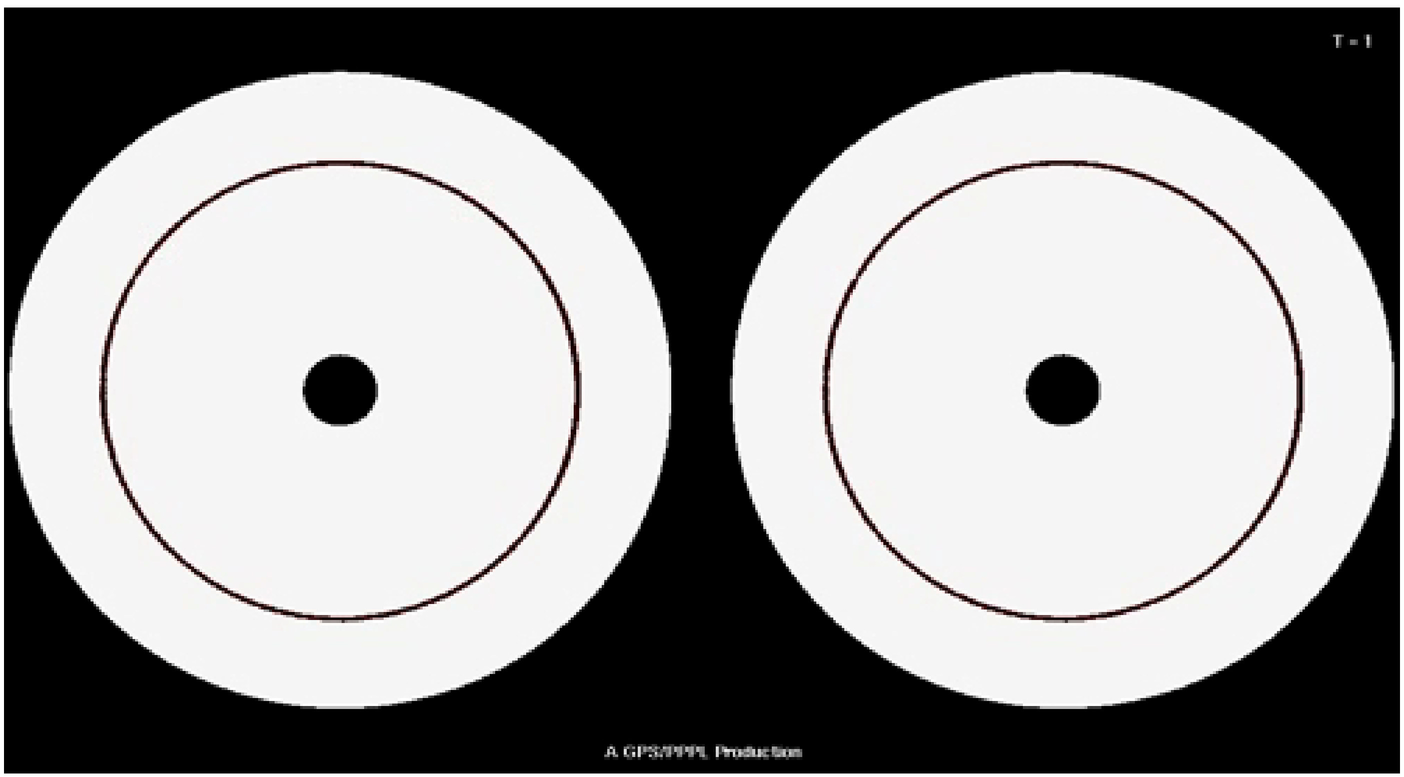
Good news for ITER!



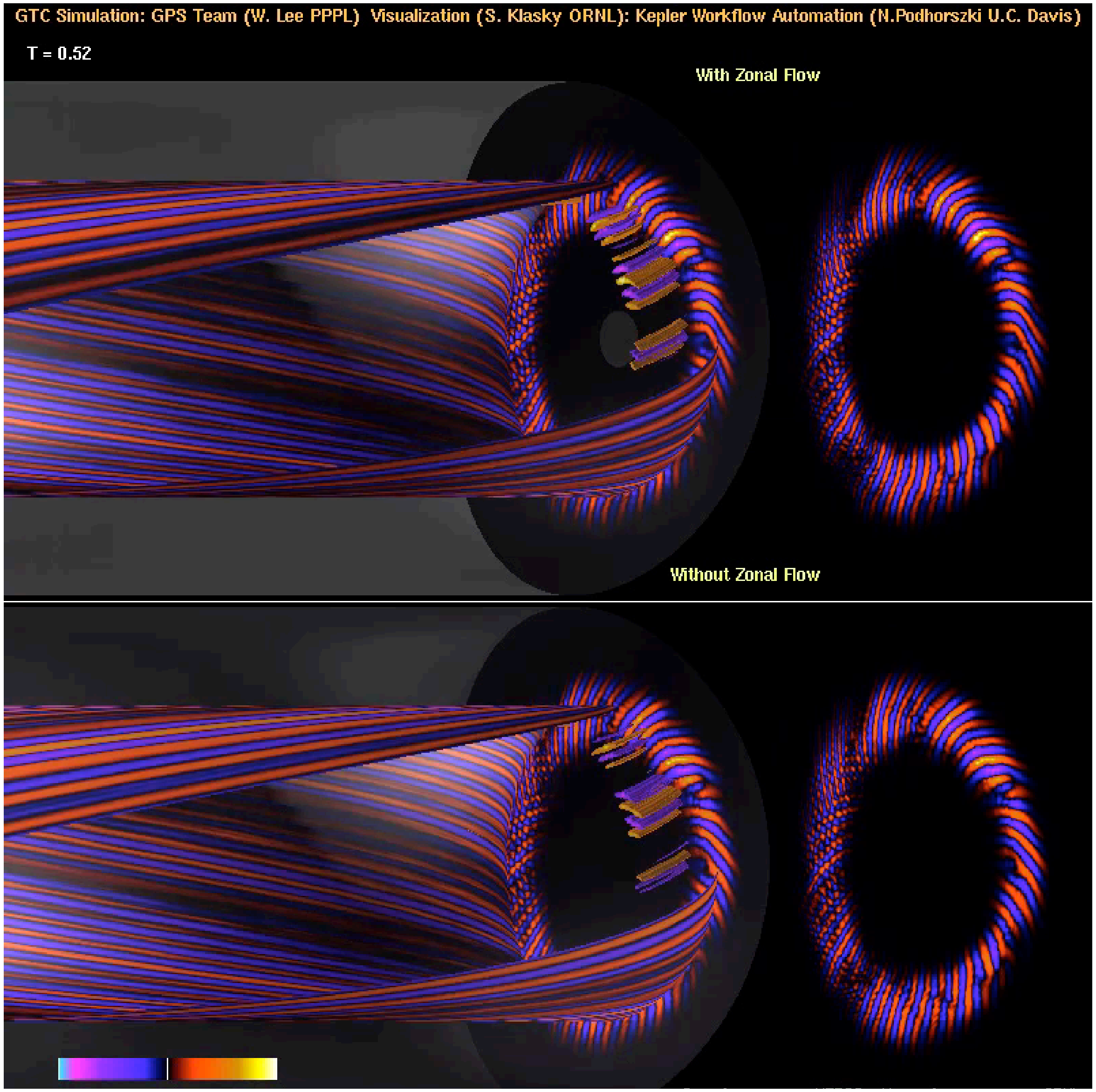


Comparison Visualization of Particle Flow

- ***Challenge:*** ***5D+time phase space visualization***
  - $4x10^9$ particles in present simulations
- ***Physics Insight:*** ***Particles diffuse faster with velocity space non-linearity included in simulations***
  - Visualization shows movement of particles around the eddies
- ***Comparison visualization*** -- ***effective tool aiding scientific discovery***

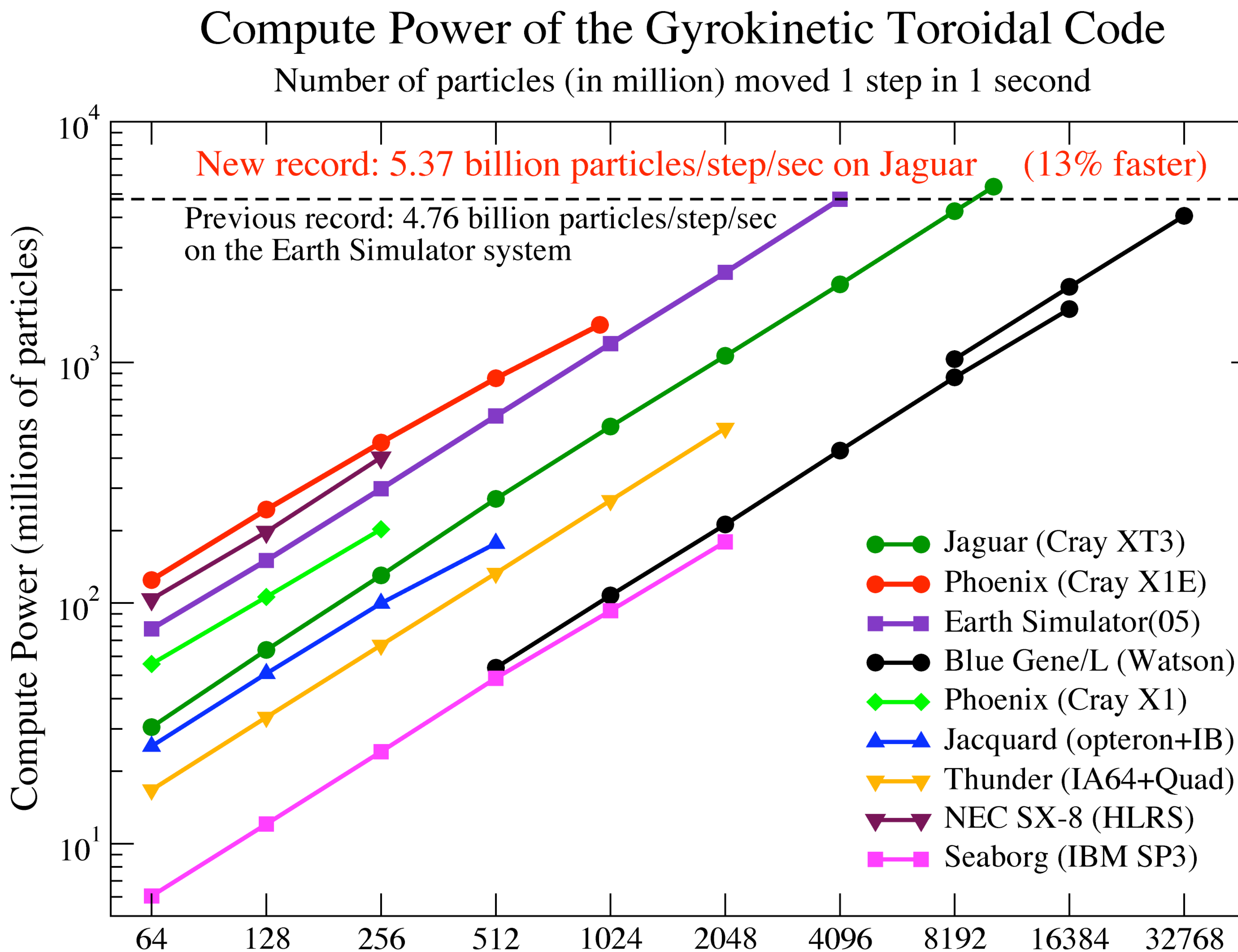


Recent High-Resolution Simulations



- Recent high-resolution visualization from ***realistic shaped-cross section toroidal plasma simulations*** on leadership class computers

- Efficiently generated via "Workflow Automation" -- ***automation of data movement, data reduction, data analysis, and data visualization***
[SciDAC SDM Center's Kepler workflow project (S. Klasky, et al.)]

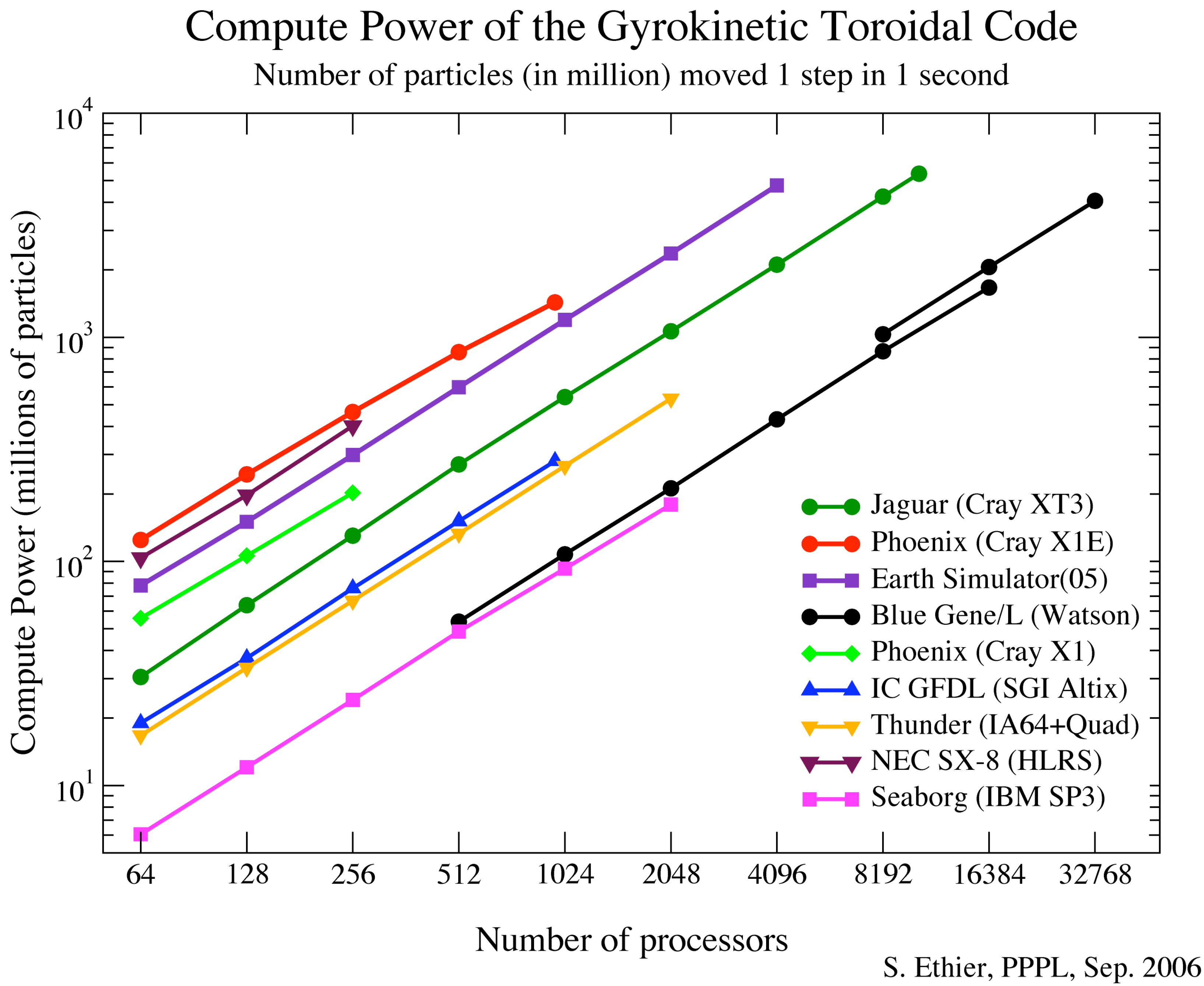


Compute Power of the Gyrokinetic Toroidal Code
Number of particles (in million) moved 1 step in 1 second
New record: 5.37 billion particles/step/sec on Jaguar (13% faster)
Previous record: 4.76 billion particles/step/sec
on the Earth Simulator system
Compute Power (millions of particles)
10^4
10^3
10^2
10^1
64
128
256
512
1024
2048
4096
8192
16384
32768
Jaguar (Cray XT3)
Phoenix (Cray X1E)
Earth Simulator(05)
Blue Gene/L (Watson)
Phoenix (Cray X1)
Jacquard (opteron+IB)
Thunder (IA64+Quad)
NEC SX-8 (HLRS)
Seaborg (IBM SP3)

Latest Results on the Leadership Class Supercomputers

- **GTC successfully utilized 4096 processors on the *Earth Simulator* vector supercomputer in Japan and achieved an unprecedented 7.2 Teraflops sustained performance**
- **GTC has succeeded in running on 10,368 processors at ORNL's Leadership Computing Facility *"Jaguar" CRAY XT3* with over 95% efficiency on the second core, advancing 5.4 billion particles per step per second to demonstrate extremely high resolution simulation within a reasonable run time**
- **After demonstrating excellent scaling on a single rack (2048 processors) IBM BGL System at ANL, GTC has now scaled to 16 racks (32,768 processors) at *IBM Blue Gene Watson* with over 90% efficiency on the second core**

- ***High resolution calculations with very low noise levels enables studies of most challenging scientific questions including nature of long-time temporal evolution of turbulence in fusion plasmas***

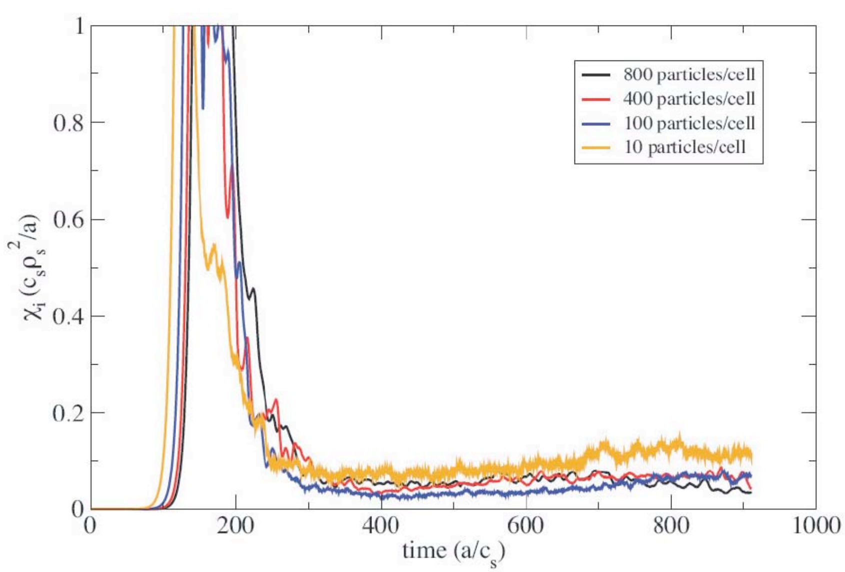


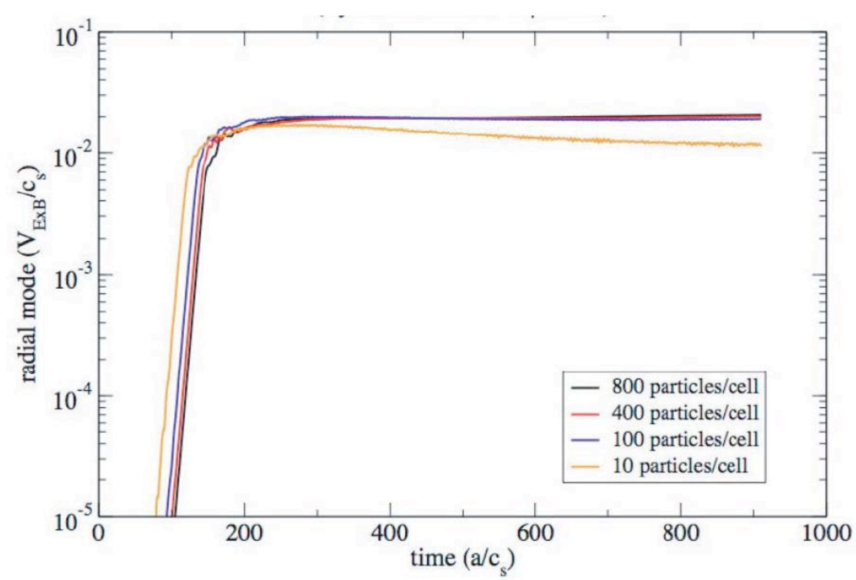
Compute Power of the Gyrokinetic Toroidal Code
Number of particles (in million) moved 1 step in 1 second
Compute Power (millions of particles)
10^4
10^3
10^2
10^1
64
128
256
512
1024
2048
4096
8192
16384
32768
Number of processors
Jaguar (Cray XT3)
Phoenix (Cray X1E)
Earth Simulator(05)
Blue Gene/L (Watson)
Phoenix (Cray X1)
IC GFDL (SGI Altix)
Thunder (IA64+Quad)
NEC SX-8 (HLRS)
Seaborg (IBM SP3)
S. Ethier, PPPL, Sep. 2006

PIC Noise Resolution in ITG Simulations

- **ITG convergence test using GTC on the CRAY X1E at ORNL: using 10, 100, 200, 800 particles per cell**

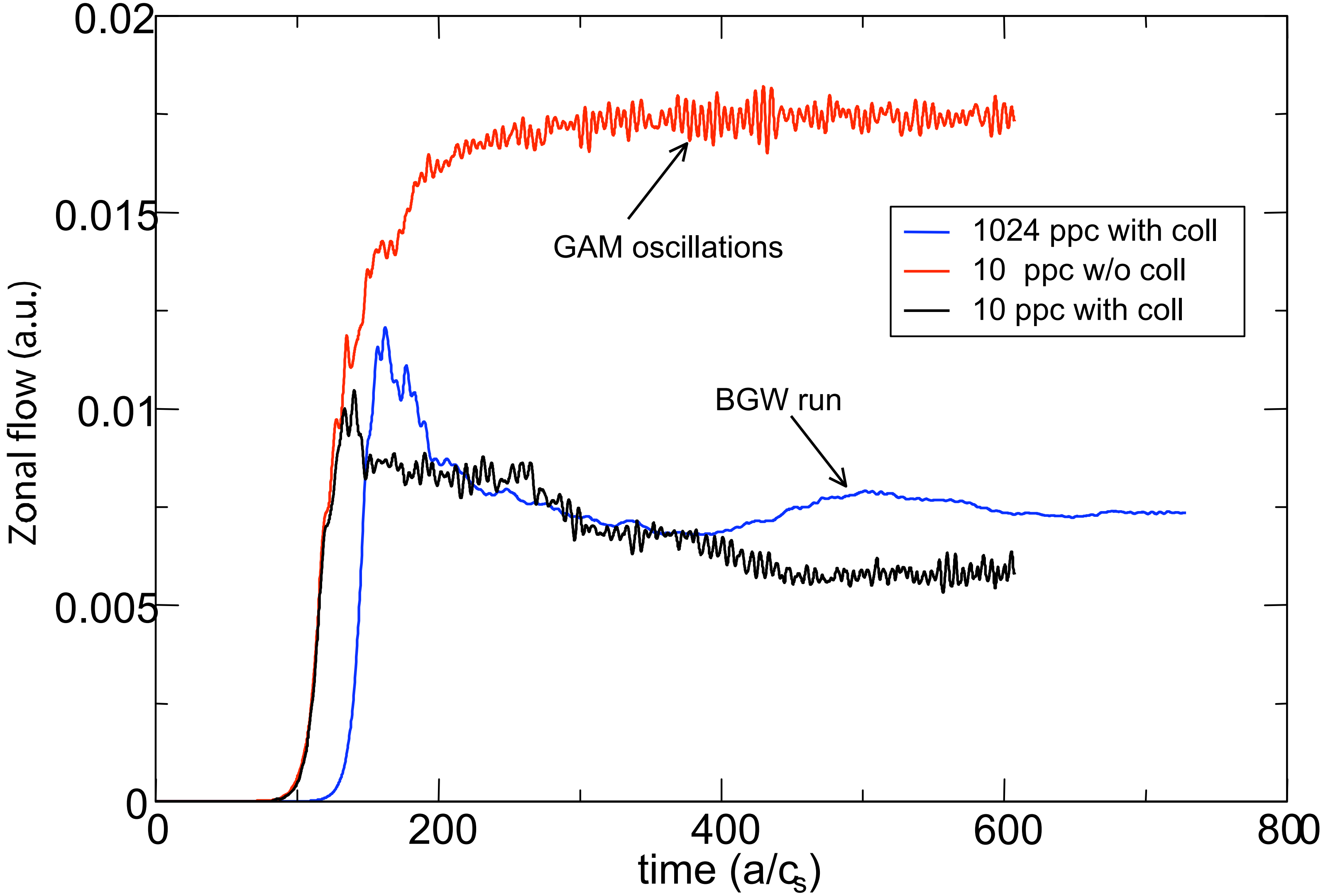
- **Results show that the saturated value of the thermal diffusivity (*left fig.*) as well as the zonal flow amplitude (*right fig.*) are virtually constant for > 10 particles per cell**



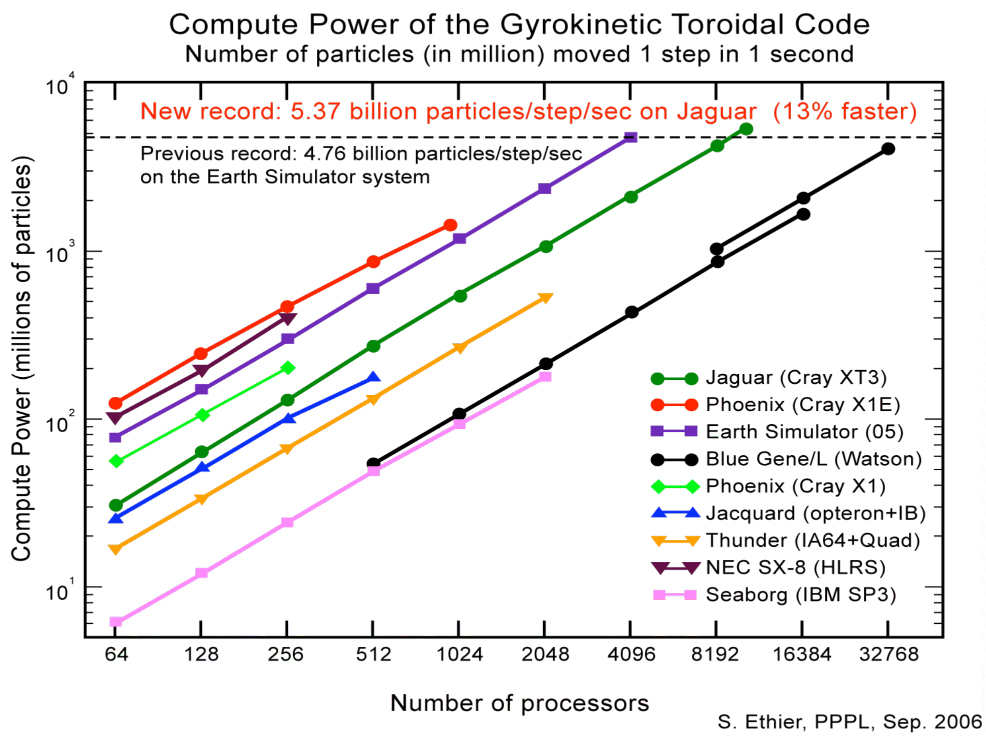


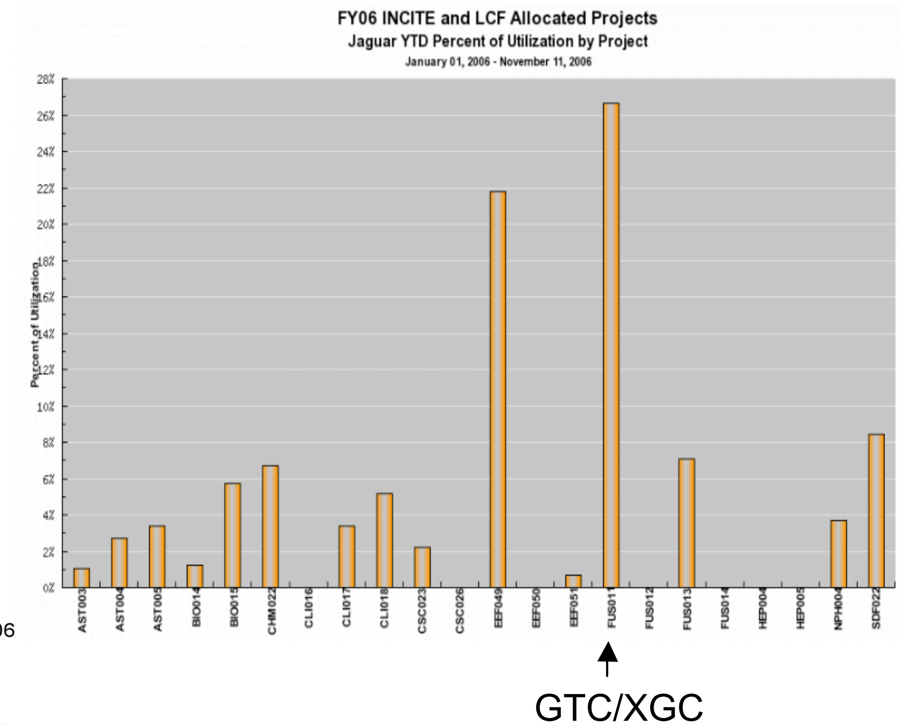
Time evolution of radial mode (zonal flow)

(Cyclone base case -- a/ρ_i =125, t_{eff}= 0.01t_{ii}, Lorentz model)

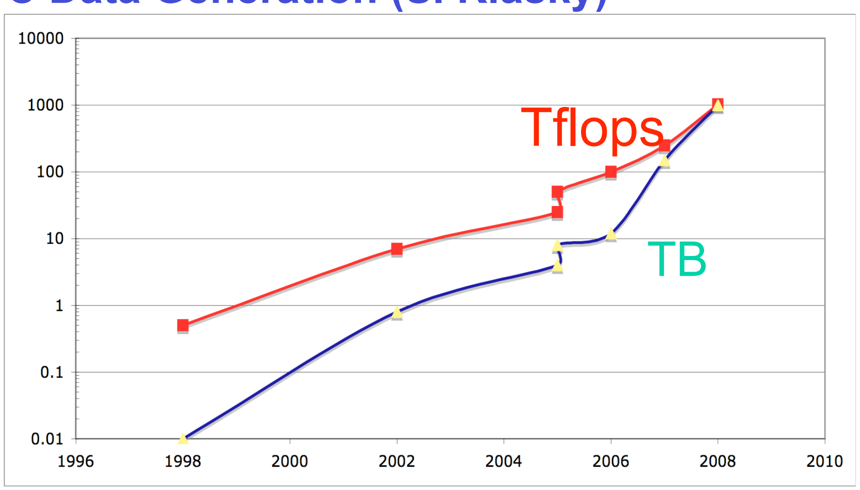


GTC Code Performance





GTC Data Generation (S. Klasky)



Increased Output due to:

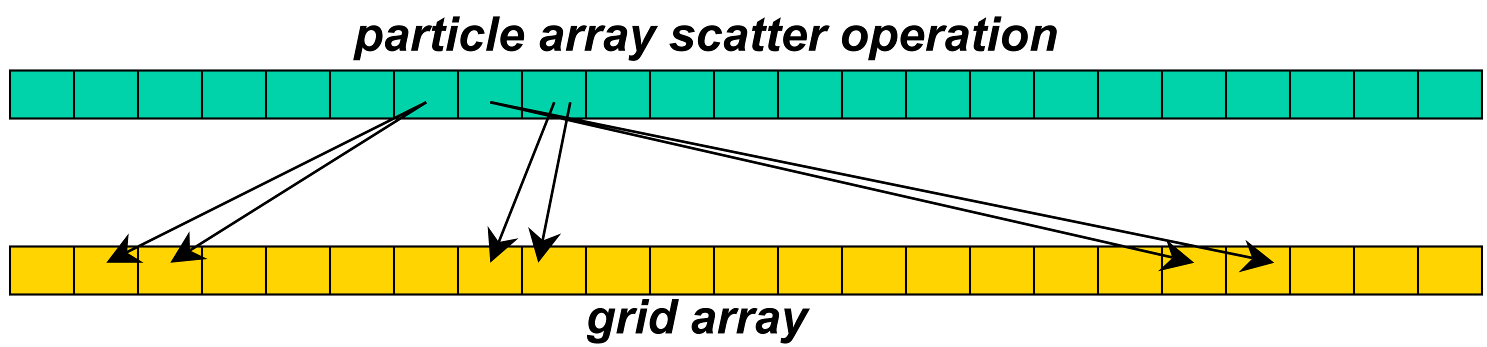
- Asynchronous metadata rich I/O
- Workflow automation
- More analysis services in the workflow

Computational Challenges

- ***Fast and Efficient Elliptic (Poisson) Solvers:***
  - Required for both Particle-in-Cell (PIC) kinetic codes and Magneto-hydrodynamics (MHD) fluid codes.
    - PIC applications involve extremely large sparse matrix system (10^8 X 10^8 grid points)
  - Deal with non-Cartesian irregular grid in toroidal geometry.
  - Need efficient pre-conditioner to speed-up the solve (e.g., pre-arranging matrix)
  - Portable parallel solver
- ***Optimization of Parallel Algorithms:***
  - Improve scalability and efficient utilization of increasing numbers of processors (fluid and kinetic codes)
  - Properly distribute particles over simulation domain in PIC simulations
  - Improve load balancing

Computational Challenges

- ***<u>"Gather-Scatter" operation in PIC codes</u>***
  – The particles are randomly distributed in the simulation volume (grid).
  – Particle charge deposition on the grid leads to indirect addressing in memory (see below).
    - need to arrange data to enable "direct-addressing" (at least for some time period)
    - also a problem in computer games
  – Not cache friendly.
  – Need to be tuned differently depending on the architecture.



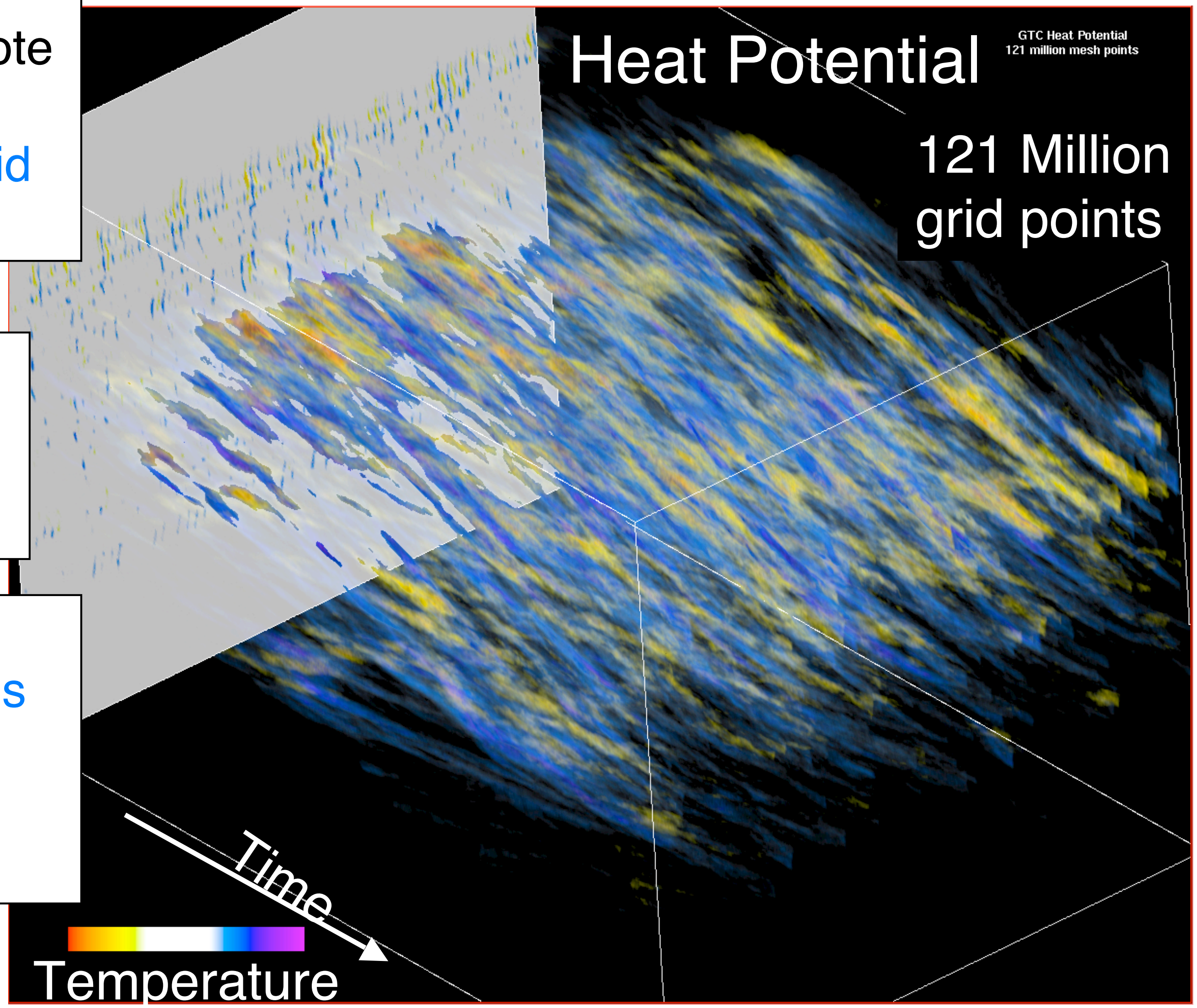
Data Analysis, Management, & Visualization Challenges

Terabytes of data are now generated at remote location (Data Management, Data Grid technologies)

Data must be efficiently analyzed to compute derived quantities

New advanced visualization techniques are needed to help identify key features in the data

Particle in Cell Turbulence Simulation



Data Analysis, Management, & Visualization Challenges

- *Data-management challenge in some scientific areas already exceeding compute-power challenge in needed resources*

- *Automated Workflow Environment:*
  - *Tera- to Peta-bytes of data to be moved automatically from simulations to analysis codes*
  - *Feature Detection/Tracking to harvest scientific information -- impossible to understand without new data mining techniques*

- *Parallel I/O Development and Support - define portable, efficient standard with interoperability between parallel and non-parallel I/O*
  - *Massively parallel I/O systems needed since storage capacity growing faster than bandwidth and access times*

- *Real-time visualization to enable "steering" of long-running simulations*

The Evolution of Science

Observational Science

- Scientist gathers data by direct observation
- Scientist analyzes data

Analytical Science

- Scientist builds analytical model
- Makes predictions.

Computational Science

- Simulate analytical model
- Validate model and makes predictions

Data Exploration Science

- Data captured by instruments and/or generated by simulations
- Processed by software
- Placed in a database / files
- Scientist analyzes database / files

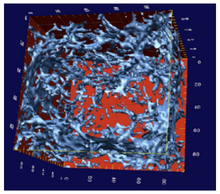
(Courtesy Jim Gray)



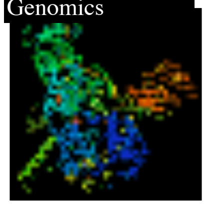
Driving Simulation Applications

Science/Engineering

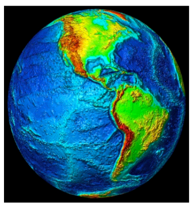
Astrophysics



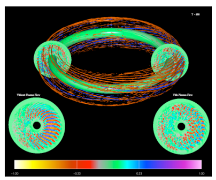
Biology, Genomics



Geosciences

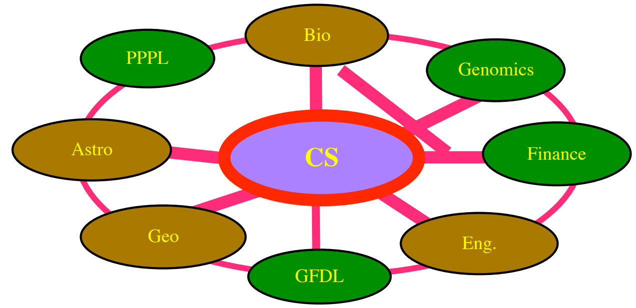


Plasma Physics



Princeton University's PICASso Program

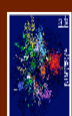
(Funded by multi-year IGERT grant from NSF)



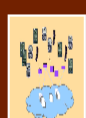
Program in Integrative Computer and Application Sciences

Scalable Services

Internet Services



Mobile Services



Information Archives

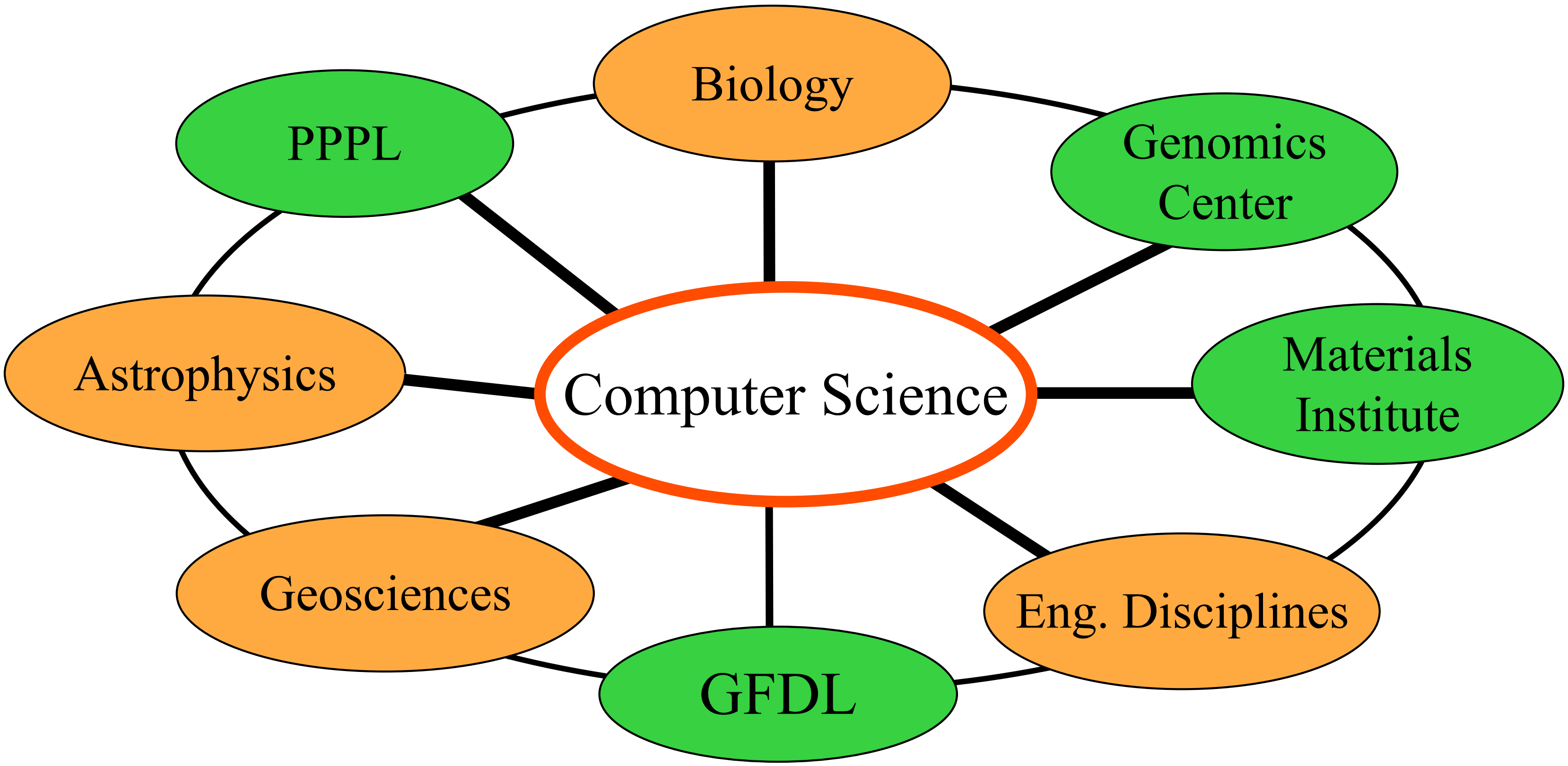
The Computational Pipeline: *Analysis Tools*

Models | Algorithms | Scalable Systems | Data Analysis & Management | Visualization

- ***Provide integrated research & training in the entire computational pipeline***
- ***Promote interdisciplinary research in computational and information science***
- ***Train a new breed of truly interdisciplinary researcher***

PICASso

"Hub-and-spoke" model centered in Computer Science



PICASso

Full Academic Courses

- COS 590: Computational Methods and their Applications Across Disciplines
- COS 598D: Data Analysis and Modeling in Science, Engineering, and Information Services
- COS 597C: Scalable Systems and Applications
- AST 302: Computational Astrophysics
- CHE 432: Dynamics of Cellular Processes

Conclusions

• **Advanced Computations provides *natural bridge* for fruitful collaborations between CS, Applied Math, & other Physical Science Applications areas such as Plasma Physics**

• **Particle-in-cell (PIC) is a very powerful method to study plasma micro-turbulence -- a key area of fusion research demonstrating "leadership class computing" impact**
- "Gather/Scatter" operations continue to pose challenge to all types of processors
- PIC holds great promise for accelerating pace of achievement of important new scientific results -- exciting advances for *time to solution* for a given resolution/accuracy
- HPC-enabled high resolution PIC simulations has begun to allow studies of most challenging scientific questions including nature of *long-time temporal evolution of turbulence in plasmas*

• **Interdisciplinary Computational Science is helping to *attract, educate, & retain young talent* essential for the future.**